RESEARCH ARTICLE

BioDiscovery

OPEN ACCESS

Lifestyle Characteristics and Dietary impact on Plasma Concentrations of Beta-carotene and Retinol

Rabindra Nath Das[1], Pradip K. Sarkar[2]

[1] Department of Statistic, The University of Burdwan, Golapbag, Rajbati, Burdwan, West Bengal 713104, India. E-mail: rabin.bwn@gmail.com
[2] Department of Basic Sciences, Center of Academics, Parker University, 2500 Walnut Hill Lane, USA. E-mail: psarkar@parker.edu

Abstract

Mechanisms of dietary micronutrients and personal characteristics of the human body are intricately complicated. These mechanisms, however, can be easily interpreted through appropriate mathematical relationships. The present study aims to detect the statistically significant impact of personal characteristics and diet on plasma concentrations of retinol and beta-carotene using statistical modeling. The present analyses indicate that age, sex, smoking habit, quetelet, vitamin use, consumed calories, fiber, and dietary beta-carotene are statistically significant factors on plasma beta-carotene levels. On the other hand age, sex, smoking status, consumed fat, and dietary beta-carotene are significant factors on plasma retinol. These analyses indicate that changes in the variances of plasma beta-carotene and retinol are non-constant. Impacts of personal characteristics and dietary factors on human plasma concentrations of retinol and beta-carotene are explained based on mathematical relationships. These analyses support many earlier researches findings. However, the analyses also identify many additional casual factors that explain the means and variances of plasma beta-carotene and retinol, which earlier researches have not reported.

Citation: Das RN, Sarkar PK. Lifestyle Characteristics and Dietary impact on Plasma Concentrations of Beta-carotene and Retinol. Biodiscovery 2012; 3: 3; DOI: 10.7750/BioDiscovery.2012.3.3

Copyright: © 2012 Das et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, provided the original authors and source are credited.

Received: 31 July 2012; **Revised:** 27 September; **Accepted:** 30 September 2012; **Available online/Published:** 30 September 2012

Keywords: cancer; joint generalized linear models; log-normal model; non-constant variance; plasma beta-carotene; plasma retinol.

*Corresponding Author: Pradip K. Sarkar, e-mail: psarkar@parker.edu

Conflict of Interests: No potential conflict of interest was disclosed and there is not any competing financial interest in relation to the work described.

Introduction

Epidemiological research often seeks to identify causal relationships between risk factors and diseases. It is well known that low human plasma concentrations of retinol, beta-carotene, or other carotenoids are strongly associated with an increased risk of developing cancer [1-4]. However, lower retinol levels may be a consequence of rather than a cause of invasive cancer [2].Some studies have been conducted to detect the determinants of human plasma concentrations of beta-carotene and retinol levels [1, 5]. A few earlier researches have reported that some dietary factors and personal characteristics are highly associated with plasma carotene levels [6-8]. Higher dietary intake of green and yellow leafy vegetables, for example, tend to increase plasma beta-carotene levels [9-11]. Compared to men, women have been reported to have higher plasma levels of retinol, beta-carotene, and other carotenoids [12-13]. Supplemental vitamin users were found to have higher levels [3-4, 14], while those who smoke cigarettes and consume alcohol have been reported to have lower beta-plasma concentrations [12, 15, 16]. Cancer researchers have aimed to identify the determinants of plasma concentrations of carotenoids, as low levels are significantly related to the development of cancer. Many researches have focused on the relationships

between beta-carotene levels, age, and obesity [3,12,14].

Many of the relationships researchers have sought to identify between carotenoids and diseases are still unclear and inconclusive. The reason is that evidences are insufficient or conflicting. Generally, validated relationships are established based on statistical analysis. Some previously reported statistical analyses indicate that certain relationships between carotenoids and disease are inconsistent. For a better understanding of these relationships, further studies are indispensable. The functional relationship is considered a probabilistic (regression or generalized linear model (GLM)) model that provides an approximation to relatively more complex phenomenon [17-20]. If the univariate response data are independent or dependent, heteroscedastic (non-constant variance) and belong to exponential family, both the mean and variance need to be modeled simultaneously, using link functions for natural mean and variance. This modeling approach is known as joint generalized linear model (JGLM) [21].

Generally, continuous positive observations belong to an exponential distribution, and their variances may or may not be constant, as the observations have variance-to-mean relationships. The problem of non-constant variance (for the response variable y) in linear regression is a departure from the standard least squares assumptions. This problem of inequality of variance occurs often in practice, frequently in conjunction with a non-normal response variable. To minimize the problem, an appropriate method is to transform the response variable to stabilize variance. This makes the distribution of the response variable closer to the normal distribution, and it improves the fit of the model to the data. However, in practice, the proper transformation may not always stabilize the variance [20, 22]. Thus, for analysis of positive data with non-constant variance, it is crucial to use joint generalized linear models (JGLMs) (modeling of mean and variance simultaneously) to identify the significant factors of the process [21, 22]. Joint GLM (along with their relevant references) for log-normal models are described in the materials and methods section.

Nierenberg et al. [5] studied the personal characteristics and dietary effects on plasma beta-carotene and retinol concentrations based on the data described in results section. To identify the appropriate model, the investigators used many statistical techniques, namely, multiple regression analysis, least squares method, multicollinearity checking tools, outliers detection tools, variance stabilization transformation, model selection criteria, and others. Nierenberg et al. [5] also noticed that the variance of the response (plasma beta-carotene concentration) was non-constant, and its distribution was non-normal. Therefore, the investigators used logarithm transformation of the plasma beta-carotene concentrations to stabilize variance and to reduce the distribution close to normal. Final data analyses have been done using log-transformation by the least squares method. Unfortunately, model fit (or index of fit) criteria measures multiple correlation (R^2) and adjusted multiple correlation (R^2_{adj}) are very small for the three derived fitted models. Specially, the maximum value of (R^2) and (R^2_{adj}) for the three derived fitted models are 0.2714 and 0.2466, respectively. These values clearly indicate that the three derived fitted models represent only some weak relationships, which may be improved further.

For heteroscedastic data, log-transformation is often recommended for stabilizing the variance [23]. In practice, though, the variance is not always stabilized by this method [20]. For example, Myers et al. [20] analyzed "The Worsted Yarn Data" using a usual (errors are uncorrelated and homoscedastic) second order response surface design. Myers et al. [20] treated the response (y=T) as the cycles to failure (T), and also noticed that the variance was non-constant and the analysis was incorrect. Then using log transformation of the cycles to failure (i.e., $y=\ln T$), the final data analysis had been done, and it was found that log model, overall, was an improvement over the original quadratic fit. The researchers noticed, however, that there was still some indication of inequality of variance. Recently, Das and Lee [22] showed that simple log transformation was insufficient to reduce the variance constant, and the investigators analyzed the data using joint generalized linear models. This study found that many factors were significant and that log-normal distribution was more appropriate. For non-constant variance of response, classical regression technique gives inefficient analysis, often resulting in an error so that significant factors are classified as insignificant. For instance, the analysis by Myers et al. [20] missed many important factors of the process. "This error is serious in any data analysis". The present authors notice that the original data set is positive, variance of the response is non-constant, distribution is non-normal, and model fit criteria measures are very small. These observations motivated us to take up this present study.

In medical research, it is very important to derive the relationship between causal factors and the disease. In statistical literature, models are mainly focused on the mean. The modeling of the dispersion has often been neglected. Analysis based on the constant variance assumption when, in fact, variance is non-constant can give inefficient analysis of the mean, often resulting in an error so that significant factors are classified as insignificant. For example, the data analysis by Nierenberg et al. [5] missed many important factors. This is very serious in medical treatments because a

wrong selection of causal factors may risk patients' lives. Therefore, it is crucial to use the appropriate statistical method to identify significant factors for deriving relationships. This article uses the joint modeling of mean and dispersion for detecting the relationship of dietary factors and personal characteristics to human plasma carotene concentrations.

Present study analyzes the relationship of two response variables (plasma beta-carotene and retinol) to the explanatory variables of dietary factors and personal characteristics. It is identified that variances of these two response variables are non-constant. Consequently, two models are derived, one for the plasma beta-carotene and the other for plasma retinol. Present analyses identify the following: Mean plasma levels of beta-carotene is explained by the statistically significant factors age, sex, smoking status, quetelet index (weight/height^2), vitamin use status, consumed calories, and fiber intake. As age increases, plasma beta-carotene levels increases. Female sex is positively associated with plasma beta-carotene; and regular vitamin and fiber intake increase plasma beta-carotene. On the other hand, increased calorie consumption, quetelet index, and current smoking status decrease mean plasma beta-carotene levels. Variance of plasma beta-carotene is increased by increased beta-carotene consumption. It is also shown to be decreased by higher fiber intake and no regular vitamin use status. Mean plasma retinol increases with age and former smoking status, but decreases only with increased fat consumption. Plasma retinol variance increases with increased beta-carotene intake, and is decreased in females in comparison to males.

Results

A. Data: Plasma data set under the present study contains 315 observations on 14 variables. Study subjects (N=315) were patients who had an elective surgical procedure during a three-year period to biopsy or remove a lesion of the lung, colon, breast, skin, ovary, or uterus. The lesions were all found to be non-cancerous. The related reference to this data set is Nierenberg et al. [5]. Source of the data set is at (http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/DataSets?CGISESSID=10713f6d891653ddcbb7ddbdd9cffb79). This annotated S data frame was prepared by Hong Yu, a graduate student at the University of Virginia, December14, 2002.

B. Variables: Table 1 presents a description of each set of items and how they are operationalized for the present study.

1. Dependent variables: The dependent variables in the present study are the plasma beta-carotene and retinol levels (Table 1).

2. Independent variables: There are two sets of independent variables, qualitative and quantitative. Three independent variables (sex, smoking status, and vitamin use) are qualitative and the remaining nine are continuous variables.

Descriptive Statistics

This data set contains 42 (13.3%) male and 273 (86.7%) female patients. Number of subjects in the groups for never smokers, former smokers, current smokers are 157 (49.8%), 115 (36.5%), 43 (13.7%) respectively and for vitamin users (yes, fairly often), (yes, not often), (no)

Table 1. Operationalization of variables in the analysis.

Domain/Variable Name	Operationalization
AGE(x_1)	Age(years)
SEX(A)	Sex (1=Male (42 (13.3%)), 2=Female (273 (86.7%))
SMOKSTAT(B)	Smoking status (1=Never (157 (49.8%)), 2=Former (115 (36.5%)), 3= Current Smoker (43 (13.7%))
QUETELET(x_4)	Quetelet (weight/height^2)
VITUSE(C)	Vitamin Use (1=Yes, fairly often (122 (38.7%)), 2=Yes, not often (82 (26.0%)), 3=No (111 (35.3%)))
CALORIES(x_6)	Number of calories consumed per day
FAT(x_7)	Grams of fat consumed per day
FIBER(x_8)	Grams of fiber consumed per day
ALCOHOL(x_9)	Number of alcoholic drinks consumed per week
CHOLESTEROL(x_{10})	Cholesterol consumed (mg per day)
BETADIET(x_{11})	Dietary beta-carotene consumed (µgper day)
RETADIET(x_{12})	Dietary retinol consumed (µg per day)
BETAPLASMA	Plasmabeta-carotene (ng/ml)–Dependentvariable
RETAPLASMA	Plasma retinol (ng/ml)–Dependent variable

are 122 (38.7%), 82 (26.0%), 111 (35.3%) respectively (Table 1).

Table 3 shows that both the levels of beta-carotene and retinol increase with age, indicating that both the dependent variables may be positively associated with the factor age. Mean beta-carotene and retinol levels are respectively higher in females and males. Mean beta-carotene levels is the highest for "never smoking status group". The order of beta-carotene levels was as follow from highest to lowest: never > former > current smoking status. Mean retinol concentration is the highest for former smoking status, and is independent at the other two smoking groups. Tables 4 and 5 show that both the mean levels of beta-carotene and retinol decrease with quetelet and fat consumed, indicating that they may be negatively associated separately with quetelet and fat intake. Both the mean levels of beta-carotene and retinol are maximum at vitamin use (1 = yes, fairly often), and seem to be decreasing with respect to (1 = yes, fairly often), (2 = yes, not often) and (3 = no) vitamin use status (Table 4). Table 5 shows that the mean levels of beta-carotene increases, while retinol concentration decreases with fiber intake. Both the mean levels of beta-carotene and retinol increase with the increased alcohol consumption (Table 5). Mean levels of plasma beta-carotene decreases, while the mean retinol levels is indifferent with cholesterol intake (Table 6). Table 6 also shows that the mean levels of beta-carotene increases, while the mean retinol levels decreases with the beta-carotene diet intake. Beta-carotene and retinol mean concentrations decrease with the consumed retadiet and calories (Table 7). Standard deviations of both the beta-carotene and retinol concentrations change along with most of the explanatory variables, indicating that both the variances may be non-constant (Tables 3-7). Tables 3-7, show the behavior of both the dependent variables, plasma levels of beta-carotene and retinol, in relation to the independent variables.

Beta-carotene Plasma Levels Data Analysis

This subsection analyzes plasma levels of beta-carotene, treating it as the response variable, in relation to the 12 covariates (Table 1) as explanatory variables. There are three qualitative characters (factors) and nine continuous variables. For factors, the constraint

Table 2. Means and standard deviation of all variables in the analysis.

Domain/Variable Name	Mean	Standard deviation (s.d.)
AGE	50.1460	14.5752
SEX	1.8667	0.3405
SMOKSTAT	1.6381	0.7110
QUETELET	26.1577	6.0136
VITUSE	1.9651	0.8607
CALORIES	1796.6546	680.3474
FAT	77.0333	33.8294
FIBER	12.7886	5.53302
ALCOHOL	3.2794	12.3229
CHOLESTEROL	242.4606	131.9916
BETADIET	2185.6032	1473.8865
RETADIET	832.7143	589.2890
BETAPLASMA	190.2127	182.7554
RETAPLASMA	602.7905	208.8955

Table 3. Mean (M.) and standard deviation (s.d.) of plasma beta-carotene (Betplas) and plasma retinol (Retplas) with respect to mean of age, and levels of sex and smoking status.

Variable	Age(≤ 50)	Age(> 50)	Sex1	Sex2	Smokst1	Smokst2	Smokst3
M.Betplas	189.5956	191.0682	148.6667	196.6044	206.6943	193.4696	121.3256
s.d.Betplas	199.5733	157.2258	133.6000	188.5704	192.6855	191.6395	78.8116
M.Retplas	564.5902	655.7500	700.7381	587.7216	583.3057	644.2435	563.0698
s.d.Retplas	186.8019	226.3884	307.8088	185.4307	187.6431	231.1676	206.5778

Table 4. Mean (M.) and standard deviation (s.d.) of plasma beta-carotene (Betplas) and plasma retinol (Retplas) with respect to mean of quetelet and levels of vitamin use.

Variable	Quet (≤ 26)	Quet (> 26)	Vitau 1	Vitau 2	Vitau 3
M.Betplas	224.6995	135.6557	241.7869	185.6585	136.8919
s.d.Betplas	215.5174	89.8490	244.4178	144.1367	92.1611
M.Retplas	606.5078	596.9098	613.1639	597.2683	595.4685
s.d.Retplas	211.4711	205.4825	223.8304	192.0211	205.2004

Table 5. Mean (M.) and standard deviation (s.d.) of plasma beta-carotene (Betplas) and plasma retinol (Retplas) with respect to mean of fat, fiber, alcohol.

Variable	Fat (≤ 77)	fat(> 77)	fiber (≤ 12)	fiber (> 12)	Alco (≤ 3.0)	Alco (> 3.0)
M.Betplas	203.6056	172.3556	164.1677	215.4437	181.1245	216.0366
s.d.Betplas	198.9673	157.5433	152.4681	205.2837	163.0144	229.0189
M.Retplas	611.6111	591.0296	618.2387	587.8250	579.8927	667.8537
s.d.Retplas	206.0704	212.8030	216.0412	201.2778	186.9810	215.4568

Table 6. Mean (M.) and standard deviation (s.d.) of plasma beta-carotene (Betplas) and plasma retinol (Retplas) with respect to mean of cholesterol and dietary beta-carotene.

Variable	Chol (≤ 242)	Chol (> 242)	Beta diet (≤ 2185)	Beta diet (> 2185)
M.Betplas	211.9681	158.0079	165.7302	226.9365
s.d.Betplas	210.5996	125.1314	134.2669	233.4960
M.Retplas	602.3032	603.5118	615.8889	583.1429
s.d.Retplas	207.9247	211.1476	208.7302	208.4268

Table 7. Mean (M.) and standard deviation (s.d.) of plasma beta-carotene (Betplas) and plasma retinol (Retplas) with respect to mean of consumed dietary retinol and calories.

Variable	Reta diet (≤ 832)	Reta diet (> 832)	Calo (≤ 1796)	Calo (> 1796)
M.Betpla	200.5319	174.9370	191.0278	189.1259
s.d.Betpla	206.8746	139.0083	178.3465	189.1385
M.Retpla	606.8883	596.7244	608.5889	595.0593
s.d.Retpla	211.3800	205.8441	209.8773	208.1065

that the effects of the first levels are zero is accepted. Therefore, it is taken that the first level of each factor as the reference level by estimating it as zero. Suppose that αi for i=1,2,3 represents the main effect of A. It is taken $\hat{\alpha}_1=0$, so that $\hat{\alpha}_2=\hat{\alpha}_2-\hat{\alpha}_1$. For example, the estimate of the effect $A2$ means the effect of difference between the second and the first levels in the main effect A, i.e., $\hat{\alpha}_2-\hat{\alpha}_1$.

The present article aims to examine the effects of different personal characteristics and dietary factors (explanatory variables) on plasma levels of beta-carotene, treated as the response variable. Thus, joint log-normal model (in materials and methods section) is fitted, and the results are displayed in Table 8. The selected models have the smallest Akaike information criterion (AIC) value in each class. It is well known that AIC selects a model which minimizes the predicted additive errors and squared error loss (Hastie et al., [24], p. 203-204). The value of AIC of the selected model (Table 2) is $3732.0+2\times15 = 3762.0$.

Figure 1(a) displays the histogram of residuals. It does not show any lack of fit for missing variables. Figure 1(b) presents the absolute residuals plot with respect to fitted values. This is a flat diagram with the running mean, indicating that variance is constant under joint GLM log-normal fitting. Figure 2(a) and Figure 2(b), respectively, display the normal probability plot for the mean and the variance model in Table 2. Neither figure

Table 8. Results for mean and dispersion models of plasma beta-carotene data from log-normal fit.

	Covariate	Estimate	s.e.	t	P-value	95%C.I.
Mean Model	Constant	5.2161	0.31122	16.760	<0.001	4.606-5.826
	AGE	0.0074	0.00277	2.688	0.01	0.002-0.013
	SEX2	0.2719	0.11486	2.367	0.02	0.047-0.497
	SMOKSTAT2	−0.1179	0.07925	−1.487	0.14	−0.273-0.037
	SMOKSTAT3	−0.2742	0.11507	−2.383	0.02	−0.500- −0.049
	QUETELET	−0.0333	0.00613	−5.438	<0.001	−0.045- −0.021
	VITUSE2	−0.0349	0.09553	−0.366	0.71	−0.222-0.152
	VITUSE3	−0.2980	0.08726	−3.414	<0.001	−0.469- −0.127
	CALORIES	−0.0001	0.00006	−1.704	0.09	−0.0002-0.000
	FIBER	0.0300	0.00710	4.230	<0.001	0.016-0.044
Dispersion Model	Constant	−0.3602	0.2361	−1.526	0.13	−0.823-0.103
	FIBER	−0.0484	0.0170	−2.840	0.01	−0.081- −0.015
	BETADIET	0.0002	0.0001	2.264	0.02	0.000-0.004
	VITUSE2	−0.3159	0.2092	−1.510	0.13	−0.725-0.094
	VITUSE3	−0.4284	0.1937	−2.211	0.03	−0.808- −0.049

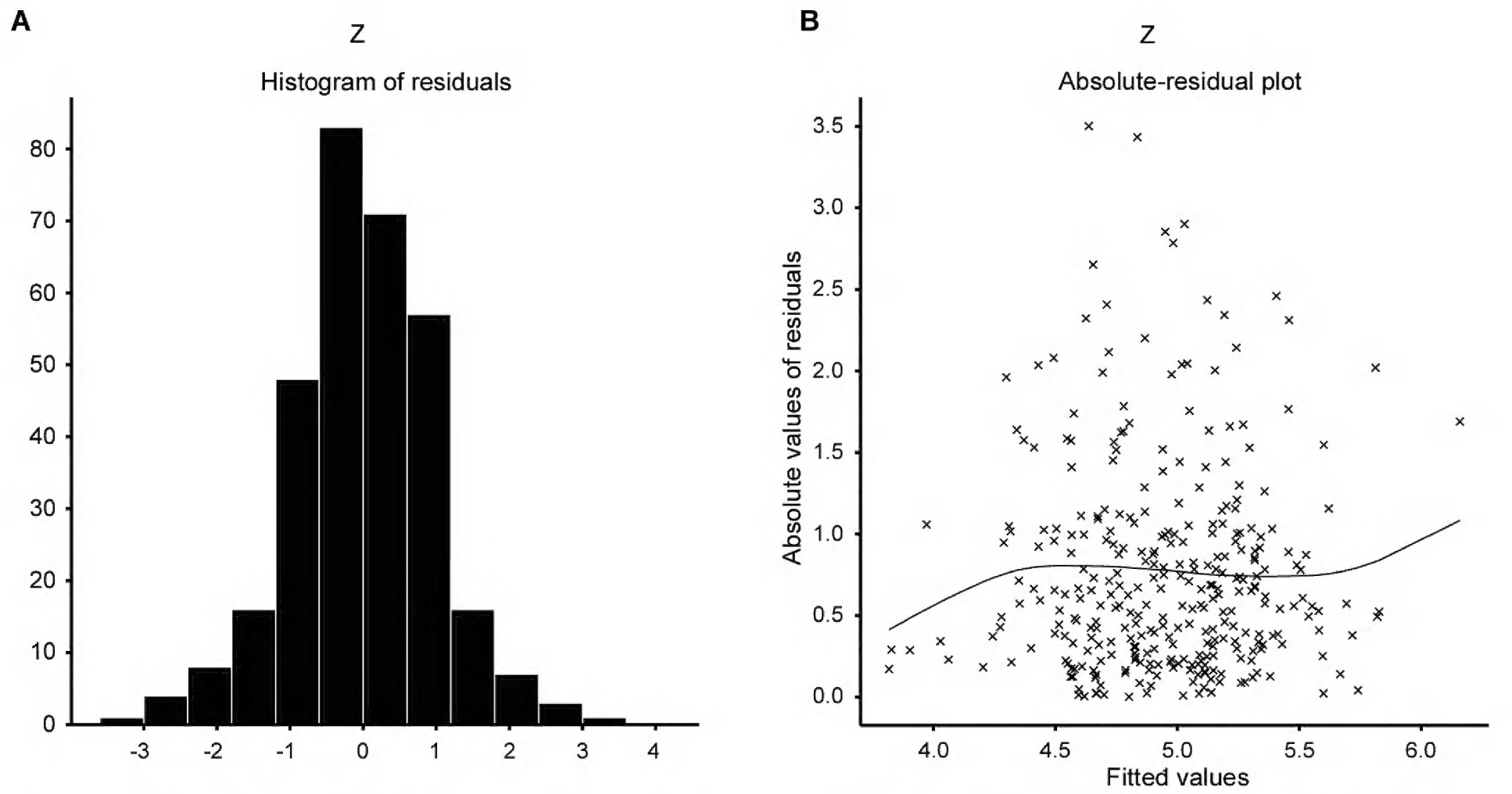

Figure 1. (a) The histogram plot of residuals and (b) the absolute residuals plot with respect to fitted values for plasma beta-carotene data (Table 2).

shows any systematic departure, indicating no lack of fit of the selected final models.

Fitted mean and variance models (Table 8) of plasma beta-carotene levels, respectively are:

$$\hat{\mu}_z = 5.2161 + 0.0074x1 + 0.2719A2 - 0.1179B2 - 0.2742B3 - 0.0333x4 - 0.0349C2 - 0.2980C3 - 0.0001x6 + 0.0300x8 \quad (1)$$

$$\hat{\sigma}_z^2 = e^{-0.0302 - 0.0484x8 + 0.0002x11 - 0.3159C2 - 0.4284C3} \quad (2)$$

Retinol Plasma Levels Data Analysis

This subsection presents the analysis of plasma levels of retinol, which is treated as the response variable, and other variables are treated as explanatory. Joint log-normal models (in materials and methods section) are fitted for the retinol data, and the results are presented in Table 9. The selected models have the smallest AIC value (4168.0 + 2 × 8 = 4184.0; Table 9) in each class.

Figure 3(a) and Figure 3(b) display the histogram

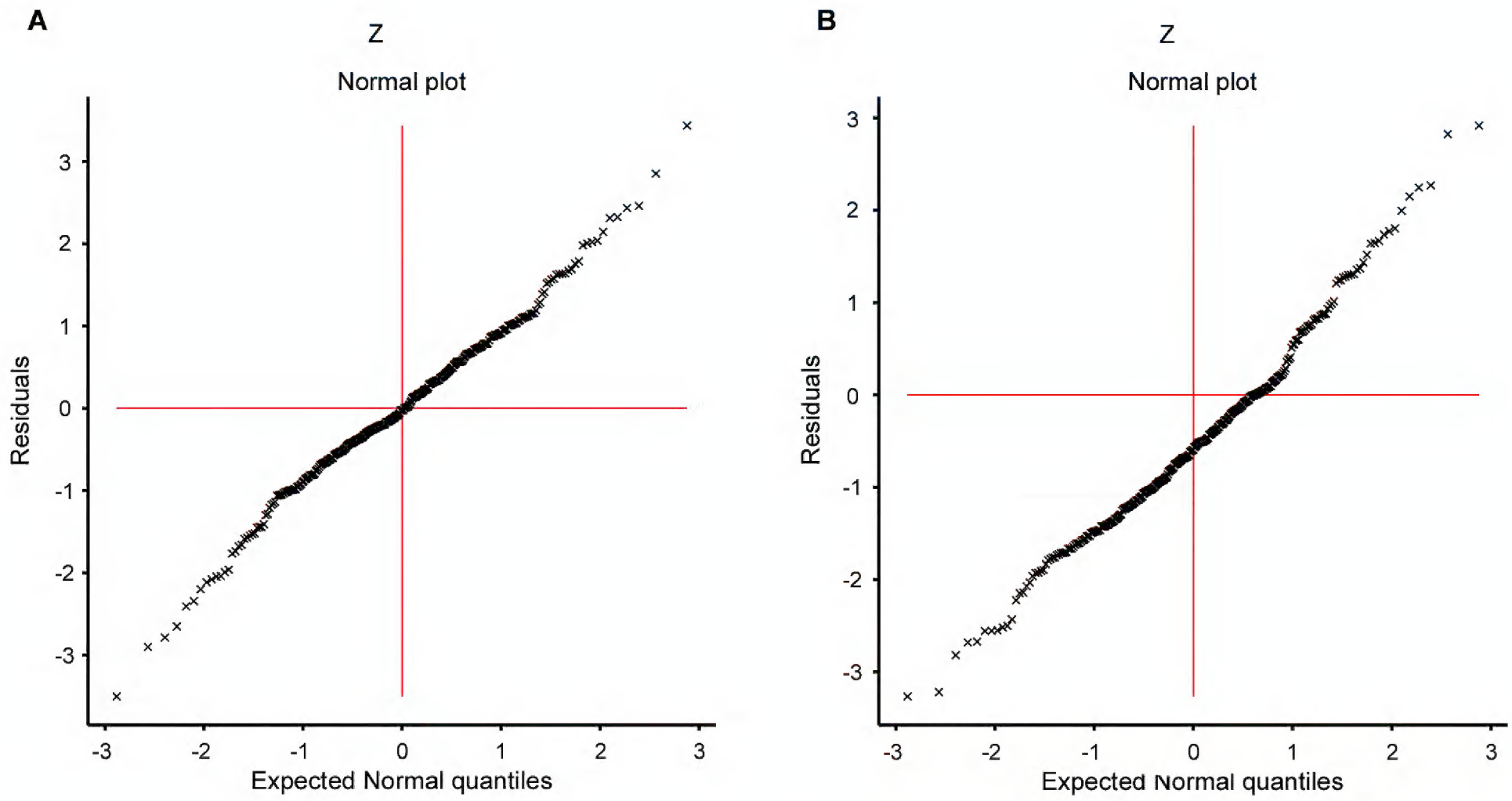

Figure 2. The normal probability plot of the (a) mean and (b) variance for plasma beta-carotene data (Table 2).

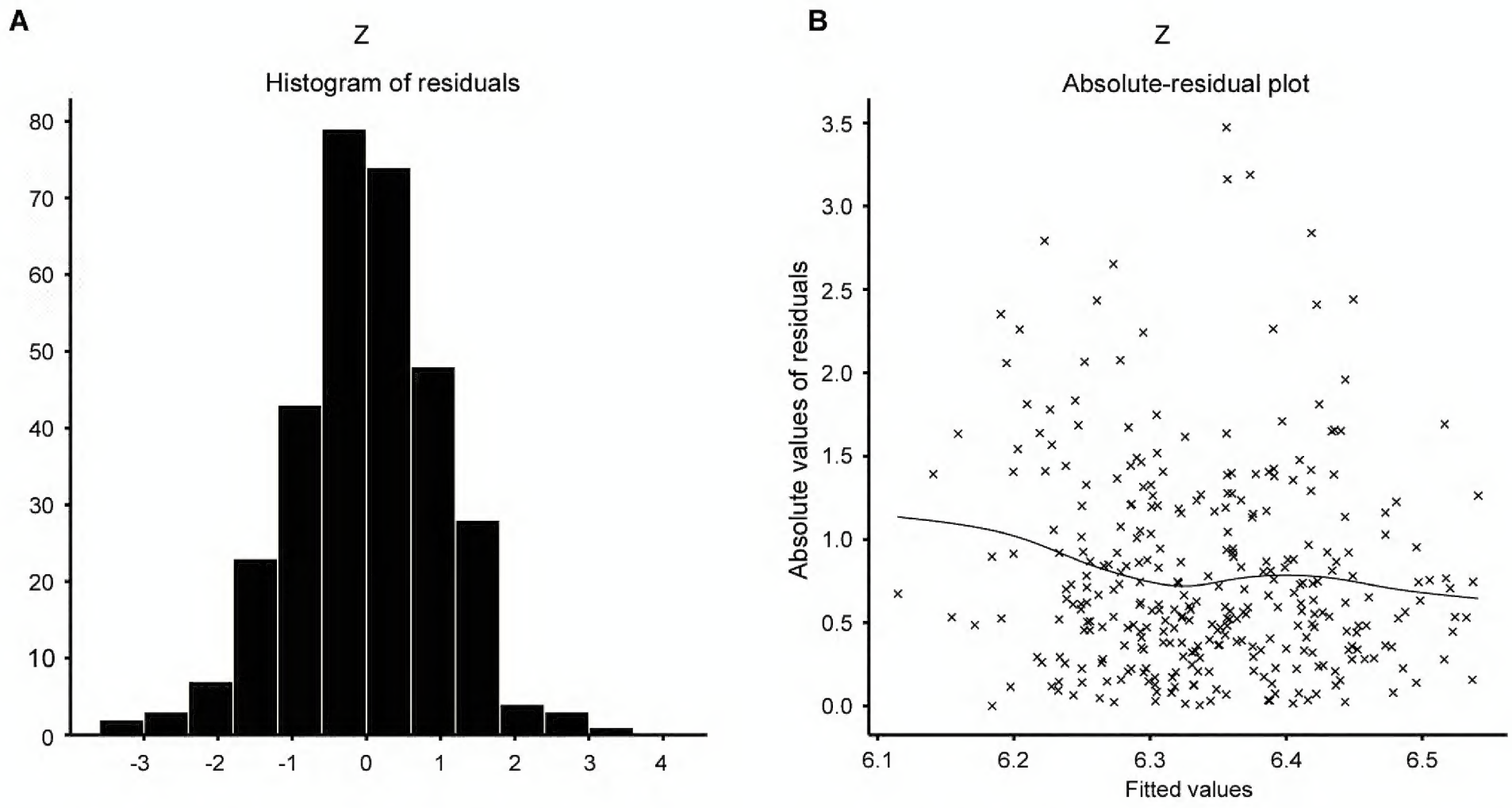

Figure 3. (a) The histogram plot of residuals and (b) the absolute residuals plot with respect to fitted values for plasma retinol data (Table 3).

of residuals and absolute residuals plot with respect to fitted values. Figure 3(a) does not show any lack of fit for missing variables. Figure 3(b) is a flat diagram with the running mean, indicating that variance is constant under the joint GLM log-normal fitting. Figure 4(a) and Figure 4(b) display respectively the normal probability plot for the mean and variance model in Table 9. Normal probability plots do not show any systematic departure, indicating no lack of fit of the selected models.

Fitted mean and variance models (Table 9) of plasma retinol levels, respectively, are:

$$\hat{\mu}_z = 6.159 + 0.004x1 + 0.080B2 + 0.003B3 - 0.001x7 \quad (3)$$

$$\hat{\sigma}_z^2 = e^{-1.9126 - 0.7379A2 + 0.0001x11} \quad (4)$$

Table 9. Results for mean and dispersion models of plasma retinol data from log-normal fit.

	Covariate	Estimate	s.e.	t	P-value	95% C.I.
Mean Model	Constant	6.159	0.08375	73.53	<0.001	5.995-6.323
	AGE	0.004	0.00126	3.53	<0.001	0.006
	SMOKSTAT2	0.080	0.03905	2.06	0.04	0.003-0.156
	SMOKSTAT3	0.003	0.05372	0.06	0.95	−0.102-0.108
	FAT	−0.001	0.00054	−1.62	0.11	−0.002-0.000
Dispersion Model	Constant	−1.9126	0.2457	−7.784	<0.001	−2.393- −1.431
	SEX2	−0.7379	0.2373	−3.109	<0.001	−1.203- −0.272
	BETADIET	0.0001	0.0001	2.399	0.02	0.000-0.003

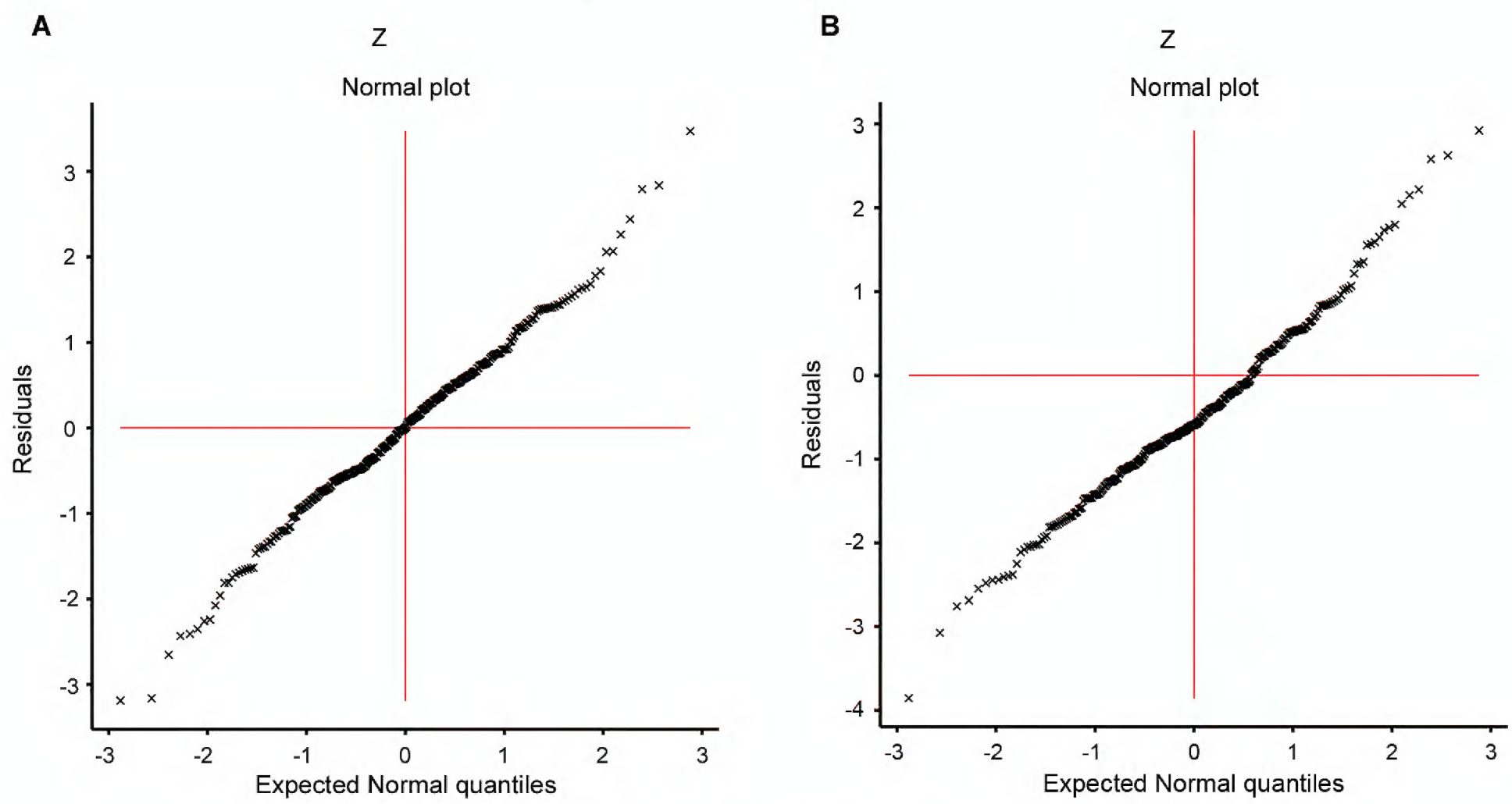

Figure 4. The normal probability plot of the (a) mean and (b) variance for plasma retinol data (Table 3).

Discussion

Table 8 (or equation 1) shows the parameters age, sex, smoking status, quetelet, vitamin use, consumed calories, and fiber intake are statistically significant (P-value ≤ 0.09) factors of mean plasma levels of beta-carotene. Mean plasma levels of beta-carotene increases with age, consumed fiber intake, regular vitamin use, and is higher in female sex, and decreases during higher calories intake, quetelet, and current smoking status. Note that smoking status (1 = never, 2 = former, and 3 = current) is negatively associated with beta-carotene. This indicates that if smoking status increases, beta-carotene decreases, and vice versa. So, beta-carotene will be minimum for maximum smoking status (i.e., 3 = current smokers). Also, vitamin use status (1 = yes, fairly often, 2 = yes, not often and 3 = no) is negatively associated with beta-carotene. In that vitamin *status* is numbered inversely to the frequency of vitamin *intake* (Table 1), this indicates that if vitamin use *status* decreases, beta-carotene increases; inversely, then, beta-carotene will be maximum for maximum vitamin intake (i.e., 1 = yes, fairly often). Mean beta-carotene is positively associated each with age and fiber consumed, and it is negatively associated each with quetelet and calories consumed. Table 8 (or equation (2)) shows that higher fiber intake, dietary beta-carotene,

and supplementary vitamin use status significantly affect the variance of plasma beta-carotene. Fiber intake is negatively, while dietary beta-carotene is positively associated with variance of beta-carotene. Thus, higher fiber intake, infrequent and no regular vitamin use, and low dietary intake of beta-carotene decrease the variance of plasma beta-carotene.

Table 9 (or equation 3) shows that age and former smoking status are directly and significantly associated with plasma retinol levels. This indicates that mean plasma retinol levels increases with age and at former smoking status. Mean plasma retinol levels is partially significant (P value = 0.11) with fat intake. The association between plasma retinol levels and fat intake is negative, indicating that plasma retinol levels decrease with increased fat consumption. Table 9 (or equation 4) shows that dietary beta-carotene is directly, while female sex is inversely associated with the variance of plasma retinol. This indicates that variance of plasma retinol levels is lower in female sex, and increases with higher intake of dietary beta-carotene.

This article focuses on the determinants of plasma levels of beta-carotene and retinol. Responses data are positive, so the probability model is log-normal or gamma [25]. Both the responses plasma beta-carotene and retinol levels are identified as non-constant variances (Tables 3–7). Thus, joint models of mean and variance are derived using log-normal distribution. The present article has examined both the joint log-normal and gamma models [22]. Observation indicates that joint log- normal models fit much better than gamma models, therefore, only the results of joint log-normal models are reported.

Tables 2–7 present the results of descriptive statistics. The variations of plasma levels of beta-carotene and retinol with respect to the explanatory variables are displayed in Tables 3–7. These results (Tables 3–7) are redundant, and also helpful to the analyzer. The main results are given in Tables 8-9; these results are supported by Tables 3–7. Tables 3–7 are displayed for better readability of the paper. These results are statistically insignificant.

Early researches pointed out that the variances of plasma levels of both the beta- carotene and retinol are non-constant [5, 26]. Those researches derived the mean model based on logarithm transformation of responses. This present study has derived both the mean and variance of plasma beta-carotene and retinol models based on joint GLM (Results Sections). Most of the present results are supported by early researches [5, 27,28]. However, some of the present results are little cited in the literature. For example, the present analysis first derived the determinants of the variances of both plasma beta-carotene and retinol (Results Section). Moreover, some additional factors were identified in the mean models (Results Section). As a result of this approach, this report attempts to remove some conflicts of earlier researches. For instance, in the literature, there are conflicting reports on the effects of alcohol, cholesterol intake, and age on plasma levels of beta-carotene. Earlier researches noted that ethanol drinkers have lower levels of plasma beta-carotene [3, 12]. However, Table 5 shows that alcohol consumers have higher levels of plasma beta-carotene. Table 10 presents the analysis of plasma levels of beta-carotene with the additional factors, alcohol and cholesterol. Analysis (Table 10) shows that alcohol and cholesterol intake are statistically insignificant, as statistically results are considered significant at a maximum of 5%. In epidemiology, partially significant factors (treated as confounders) are considered, as they may have some effects on the responses. In view point of epidemiology, alcohol intake is marginally significant (P-value = 0.19), and it is positively associated with the mean plasma beta-carotene level (supported by Table 5). Cholesterol intake (Table 10) is partially (P-value = 0.26) inversely associated with the variance of plasma beta-carotene (supported by Table 6).

Results subsections present the statistically significant determinants of plasma levels of both the beta-carotene and retinol (Tables 8, 9). For example, quetelet is inversely associated with plasma beta-carotene levels. This indicates that many obese persons have lower blood levels of plasma beta-carotene, even after adjustment for dietary intake. Obese persons have large volumes of fat stores. However, fat store is inversely (partially significant) associated with beta-carotene levels. Fat, as the partially significant factor, is not shown in Table 8, but it is close to significant (P-value = 0.11) in Table 9. Thus, plasma beta-carotene level is lower for many obese people due to their large volumes of fat stores. This conclusion is simply derived from the mathematical relationship. In view of pharmacokinetic mechanisms, however, fat may dissolve ingested vitamins. Consequently, the vitamin level will be low, indicating a low level of plasma beta-carotene.

This study found age to be directly associated with both the plasma levels of beta-carotene (Table 8) and retinol (Table 9) (supported by Table 3). Many research reports missed this factor [5]. Both the blood levels of plasma beta-carotene and retinol will increase with age. This may be due to physiological age-related changes in the human body. Moreover, this study identifies fiber and calorie intake, respectively, positively and negatively associated with plasma levels of beta-carotene (Table 8). These two findings were also missed by many earlier researches [5]. Other findings such as the relationships of carotenoids to female sex, vitamin use, quetelet, and smoking status are partially supported by earlier researches [5, 12, 27]. For example, female sex is only directly significant with the mean plasma levels of beat-

Table 10. Results for mean and dispersion models of plasma beta-carotene data from log-normal fit.

	Covariate	Estimate	s.e.	t	P-value
Mean Model	Constant	5.2534	0.31097	16.894	0.00
	AGE	0.0067	0.00276	2.424	0.02
	SEX2	0.2833	0.12846	2.205	0.03
	SMOKSTAT2	−0.1366	0.07826	−1.746	0.08
	SMOKSTAT3	−0.3014	0.11325	−2.661	0.02
	QUETELET	−0.0319	0.00601	−5.300	0.00
	VITUSE2	−0.0346	0.09389	−0.369	0.71
	VITUSE3	−0.2969	0.08727	−3.402	0.00
	CALORIES	−0.0002	0.00007	−2.369	0.02
	FIBER	0.0337	0.00733	4.597	0.00
	ALCOHOL	0.0047	0.00358	1.301	0.19
Dispersion Model	Constant	0.2220	0.3916	0.567	0.57
	FIBER	−0.0493	0.0169	−2.912	0.01
	BETADIET	0.0002	0.0001	2.388	0.02
	VITUSE2	−0.3259	0.2111	−1.544	0.12
	VITUSE3	−0.4566	0.1968	−2.320	0.02
	SEX2	−0.4658	0.2587	−1.800	0.07
	CHOLESTEROL	−0.0008	0.0007	−1.120	0.26

Table 11. Comparison between earlier and present findings.

Variable	Model	Response	Earlier findings	Present Findings
	Mean	BETAPLASMA		
AGE (x_1)	"	"	*Not identified*	+vely significant
SEX (A)	"	"	Female, +vely significant	Female, +vely significant
VITUSE (C)	"	"	1=Yes, fairly often, + vely	1=Yes, fairly often, + vely
SMOKSTAT(B)	"	"	3= Current Smoker, -- vely	3= Current Smoker, --vely
ALCOHOL(x_9)	"	"	*-- vely significant*	Insignificant
CHOLESTEROL(x_{10})	"	"	*-- vely significant*	Insignificant
FIBER(x_8)	"	"	*Not identified*	+vely significant
CALORIES(x_6)	"	"	*Not identified*	--vely significant
QUETELET(x_4)	"	"	--vely significant	--vely significant
	Variance	"		
FIBER(x_8)	"	"	*Not identified*	--vely significant
VITUSE (C)	"	"	*Not identified*	--vely significant
BETADIET(x_{11})	"	"	*Not identified*	+vely significant
	Mean	RETAPLASMA		
SEX (A)	"	"	*Female, +vely significant*	Insignificant
AGE (x_1)	"	"	*Not identified*	+vely significant
SMOKSTAT(B)	"	"	*Not identified*	2=Former, +vely significant
	Variance	"		
SEX (A)	"	"	*Not identified*	--vely significant
BETADIET(x_{11})	"	"	*Not identified*	+vely significant

carotene, but not with retinol. Current smoking appears to lower the mean plasma levels of beta-carotene (Table 8), but former (not current) smoking may increase plasma levels of retinol (Table 9). Determinants of plasma retinol levels found in this study also differ from many early researches' reports.

Finally, determinants of variances of both plasma levels of beta-carotene and retinol found in this study are completely new findings. For Beta-plasma analysis, only three factors sex, vitamin use, quetelet are identified as *confirmatory* of earlier findings. The factors age, alcohol intake, cholesterol are identified as the *conflicts* of earlier findings. The factors fiber, calories (mean model (1)) and fiber, vitamin use, beta-diet (variance model (2)) are all the new findings in the literature (Table 11). For retinol plasma analysis, all the factors age, sex, smoking status, beta-diet are completely new information in the literature (Table 11). This study may provide substantial new factors to explain the human pharmacology of both plasma levels of beta-carotene and retinol.

Materials and methods

Some continuous positive measurements in practice have non-normal error distributions, and the class of generalized linear models includes distributions useful for the analysis of such data. The problem of non-constant variance in the response variable y in linear regression is due to departure from the standard least squares assumptions. Transformation of the response variable is an appropriate method for stabilizing the variance of the response. For heteroscedastic data, the log-transformation is often recommended for stabilizing the variance [23]. However, in practice the variance may not always be stabilized despite proper transformation [20]. Box [29] proposed the use of linear models with data transformation.

For example, when

$$E(Y_i)=\mu_i \quad \text{and} \quad \text{Var}(Y_i)=\sigma_i^2\mu_i^2,$$

the transformation $Z_i=\log(Y_i)$ gives stabilization of variance $\text{Var}(Z_i)\approx\sigma^2$. However, if a parsimonious model is required, a different transformation is needed. Thus, the single data transformation may fail to meet various model assumptions. Nelder and Lee [30] proposed using joint generalized linear models (GLMs) for the mean and dispersion.

When the response Y_i is constrained to be positive log transformation $Z_i=\log Y_i$ is used. Under the log-normal distribution, a joint modeling of the mean and dispersion is such that:

$$E(Z_i)=\mu_{zi} \quad \text{and} \quad \text{Var}(Z_i)=\sigma_{zi}^2,$$
$$\mu_{zi}=x_i^t\beta \quad \text{and} \quad \log(\sigma_{zi}^2)=g_i^t\gamma,$$

where x_i^t and g_i^t are the row vectors for the regression coefficients β and γ in the mean and dispersion model, respectively. Lee and Nelder [31] studied the estimation of joint modeling of the mean and dispersion, and proposed to use the maximum likelihood (ML) estimator for the mean parameters β and the restricted maximum likelihood for the dispersion parameters γ. The restricted likelihood estimators have proper adjustment of the degrees of freedom by estimating the mean parameters, which is important in the analysis of data from quality engineering because the number of parameters of β is often relatively large compared with the total sample size [21].

Joint GLM method of estimation: Two interlinked models for the mean and the dispersion (or variance) are based on the observed data (y_i) and gamma deviance d_i, where $d_i = 2\{-\log(y_i / \hat{\mu}_j) + (y_i - \hat{\mu}_j) / \hat{\mu}_j\}$. Regression parameters are estimated by iterative weighted least squares (IWLS) method using the dispersion values which have a direct effect on the estimates of regression parameters. The whole computation is performed using two interconnected IWLS methods which are:

1. Given $\hat{\gamma}$ and the dispersion estimates, we use IWLS to update $\hat{\beta}$ for the mean model,

2. Given $\hat{\beta}$ and the estimated means, we use IWLS to update $\hat{\gamma}$ with deviances as data.

The above two steps of iteration is continued until it converges. More detailed discussions of joint generalized liner models have been described [22, 31-34].

Acknowledgement

The authors are very much indebted to referees who have provided valuable comments to improve this paper. The authors thank to Late Dr. John C. Lowe for his comments and suggestions in improving the article. The authors also thank to Mr. Hong Yu for generously providing the data set for free distribution and use for non-commercial purposes.

References

1. Hennekens CH. Micro nutrients and cancer prevention. *New Eng J Med* 1986; **315**: 1288-1289.
2. Wald NJ. Retinol, beta-carotene and cancer. *Cancer Surv* 1987; **6**: 635-651.
3. Russell-Briefel R, Bates MW, Kuller LH. The relationship of plasma carotenoids to health and biochemical factors in middle-aged men. *Am J Epidemiol* 1985; **122**: 741-749.
4. Thompson JN, Duval S, Verdier P. Investigation of carotenoids inhuman blood using high performance liquid chromatography. *J Micronutr Anal* 1985; **1**: 81-91.

5. Nierenberg DW, Stukel TA, Baron JA, Greenberg ER. Determinants of plasma levels of beta-caroteneand retinol. *Am J Epidemiology* 1989; **130**(3): 511-521.
6. Adams CF. Nutritive values of American foods. *US Department of Agriculture, Hand book Number 456*, Washington, DC: USGPO, 1975.
7. Peto R. Cancer, cholesterol, carotene, and tocopherol. *Lancet* 1981; **2**: 97-98.
8. Wald NJ, Boreham J, Hayward JL, Bulbrook RD. Plasma retinol, beta-carotene, and vitamin E levels in relation to the future risk of breast cancer. *Br J Cancer* 1984; **49**: 321-324.
9. Cornwell DG, Kruger FA, Robinson HB. Studies on the absorption of beta-carotene and the distribution of total carotenoid in human serum lipoproteins after oral administration. *J Lipid Res* 1962; **3**: 65-70.
10. Goodman DS.Vitamin Aandretinoidsinhealth anddisease. *New Eng J Med* 1984; **310**: 1023-1031.
11. Willett WC, Polk BF, Underwood BA, Stampfer MJ, Pressel S, Rosner B, et al. Relation of serum vitamins A and E and carotenoids to the risk of cancer. *New Eng J Med* 1984; **310**: 430-434.
12. Stryker WS, Kaplan LA, Stein EA, Stampfer MJ, Sober A, Willett WC. The relation of diet, cigarette smoking, and alcohol consumption to plasma beta-carotene and alpha-tocopherol levels. *Am J Epidemiol* 1988; **127**: 283-296.
13. Dimitrov NV, Boone CW, Hay MB. Plasma beta-carotene levels: kinetic patterns during administration of various doses of beta-carotene. *J Nutr Growth Cancer* 1987; **3**: 227-238.
14. Comstock GW, Menkes MS, Schober SE, Vuilleumier JP, Helsing KJ. Serum levels of retinol, beta-carotene, and alpha-tocopherol in older adults. *Am J Epidemiol* 1988; **127**: 114-123.
15. Aoki K, Ito Y, Sasaki R, Ohtani M, Hamajima N, Asano A. Smoking, alcohol drinking and serum carotenoids levels. *Jpn J Cancer Res* 1987; **78**: 1049-1056.
16. Chow CK, Thacker RR, Changchit C, Bridges RB, Rehm SR, Humble J. et al. Lower levels of vitamin C and carotenes in plasma of cigarette smokers. *J Am Coll Nutr* 1986; **5**: 305-312.
17. Chatterjee S, Price B. *Regression Analysis by Examples (3rd ed.).* New York, Wiley and Sons 2000.
18. Palta M. *Quantitative Methods in Population Health: Extensions of Ordinary Regression.*New York, Wiley and Sons 2003.
19. McCullagh P, Nelder JA. *Generalized Linear Models*. London, Chapman & Hall 1989.
20. Myers RH, Montgomery DC, Vining GG. *Generalized Linear Models with Applications in Engineering and the Sciences*. New York, John Wiley & Sons 2002.
21. Lee Y, Nelder JA, Pawitan Y. *Generalized Linear Models with Random Effects (Unified Analysis via H-likelihood)*. London, Chapman & Hall 2006.
22. Das RN, Lee Y. Log-normal versus gamma models for analyzing data from quality improvement experiments. *Quality Engineering* 2009; **21**(1): 79-87.
23. Box GEP, Cox DR. An analysis of transformations. *J Roy Stat Soc B* 1964; **26**: 211-252.
24. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. New York, Springer-Verlag 2001.
25. Firth D. Multiplicative errors: log-normal or gamma? *J Roy Stat Soc B* 1988; **50**: 266-268.
26. Khaw KT, Tazuke S, Barrett-Connor E. Cigarette smoking and levels of adrenal androgens in postmenopausal women. *New Eng J Med* 1988; **318**: 1705-1709.
27. Menkes MS, Comstock GW, Vuilleumier JP, Helsing KJ, Rider AA, Brookmeyer R. Serum beta-carotene, vitamins A and E, selenium and the risk of lung cancer. *New Eng J Med* 1986; **315**: 1250-1254.
28. Nierenberg DW, Stukel TA. Diurnal variation in plasma levels of retinol, tocopherol, and beta-carotene. *Am J Med Sci* 1987; **30**: 187-190.
29. Box GEP. Signal-to-Noise Ratios, Performance Criteria, and Transformations (with discussion). *Technometrics* 1988; **30**: 1-40.
30. Nelder JA, Lee Y. Generalized linear models for the analysis of Taguchi-type experiments. *Appl Stoch Model D A* 1991; **7**: 107-120.
31. Lee Y, Nelder JA. Generalized Linear models for the analysis of quality improvement experiments. *Can J. Stat*. 1998; **26**: 95-105.
32. Lee Y, Nelder JA. Robust Design via Generalized Linear Models. *J Qual Tech* 2003; **35**: 2-12.
33. Lesperance ML, Park S. GLMs for the analysis of robust designs with dynamic characteristics. *J Qual Tech* 2003; **35**: 253-263.
34. Qu Y, Tan M, Rybicki L. A unified approach to estimating association measures via a joint generalized linear model for paired binary data. *Commun. Stat Theory Methods* 2000; **29**: 143-156.